

Endoscopic classification of the papilla of Vater. Results of an inter- and intraobserver agreement study

United European Gastroenterology Journal
2017, Vol. 5(4) 504–510

© Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/2050640616674837

journals.sagepub.com/home/ueg



E Haraldsson^{1,2}, L Lundell^{1,3}, F Swahn^{1,4}, L Enochsson^{1,3}, JM Löhr^{1,3},
U Arnelo^{1,3} and the Scandinavian Association for Digestive Endoscopy (SADE)
Study Group of Endoscopic Retrograde Cholangio-Pancreaticography

Abstract

Background: Many endoscopists acknowledge that the appearance of the papilla of Vater seems to affect biliary cannulation. To assess the association between the macroscopic appearance of the papilla and biliary cannulation and other related clinical issues, a system is needed to define the appearance of the papilla.

Objective: The purpose of this study was to validate an endoscopic classification of the papilla of Vater by assessing the interobserver and intraobserver agreements among endoscopist with varying experience.

Methods: An endoscopic classification, based on pictures captured from 140 different papillae, containing four types of papillae was proposed. The four types are (a) Type 1: regular papilla, no distinctive features, 'classic appearance'; (b) Type 2: small papilla, often flat, with a diameter ≤ 3 mm (approximately 9 Fr); (c) Type 3: protruding or pendulous papilla, a papilla that is standing out, protruding or bulging into the duodenal lumen or sometimes hanging down, pendulous with the orifice oriented caudally; and (d) Type 4: creased or ridged papilla, where the ductal mucosa seems to extend distally, rather out of the papillary orifice, either on a ridge or in a crease. To assess the level of interobserver agreement, a web-based survey was sent out to 18 endoscopists, containing 50 sets of still images of the papilla, distributed between the four different types. Three months later a follow-up survey, with images from the first survey was sent to the same endoscopists.

Results: Interobserver agreement was substantial ($\kappa = 0.62$, 95% confidence interval (CI) 0.58–0.65) and were similar for both experts and non-experts. The intraobserver agreement assessed with the second survey was also substantial ($\kappa = 0.66$, 95% CI 0.59–0.72).

Conclusion: The proposed endoscopic classification of the papilla of Vater seems to be easy to use, irrespective of the level of experience of the endoscopist. It carries a substantial inter- and intraobserver agreement and now the clinical relevance of the four different papilla types awaits to be determined.

Keywords

Duodenoscopy, endoscopic retrograde cholangio-pancreaticography, endoscopic classification, interobserver agreement, intraobserver agreement, papilla of Vater

Received: 13 June 2016; accepted: 25 September 2016

Introduction

In the early days of fibre-optic endoscopy there were limited options for image documentation and storage. With the digital revolution and computerized image documentation, there are few, if any limitations in acquisition and storage of high-quality images during gastrointestinal endoscopy.

The ambition to standardize terminology and reporting of endoscopic procedures was emphasized

¹Department of Clinical Sciences, Intervention and Technology (CLINTEC), Karolinska Institutet, Stockholm, Sweden

²Department of Surgery, Skaraborg Hospital, Skövde, Sweden

³Center for Digestive Diseases, Karolinska University Hospital, Stockholm, Sweden

⁴Department of Surgery, Skåne University Hospital at Lund, Lund University, Lund, Sweden

Corresponding author:

Erik Haraldsson, Center for Digestive Diseases, K53, Karolinska University Hospital, Huddinge, SE-114 86 Stockholm, Sweden.

Email: erik.haraldsson@ki.se

by the World Organization of Gastrointestinal Endoscopy (OMED) in the development of the Minimal Standard Terminology.¹ The recent updated guidelines² emphasized that images, and even video sequences, captured during endoscopic examinations and therapeutic interventions are considered almost mandatory to achieve a high level of quality documentation. With the increased use of advanced imaging technologies such as high definition white-light endoscopy, conventional or virtual chromoendoscopy, probe-based confocal laser endomicroscopy,³ the demands for image storage, interpretation and standardization have expanded even further.^{4,5} Studies on the agreement between investigators in the assessment of endoscopic images are of vital importance not only for clinical practice but even more so in the design of and reporting on research protocols.

In the field of gastrointestinal endoscopy, several previous studies have dealt with endoscopists lack of concordance in image interpretation regarding specific pathological conditions. These issues have been addressed in oesophageal varices,⁶ Barrett's oesophagus,^{7,8} oesophagitis,^{9,10} eosinophilic oesophagitis,^{11,12} bleeding peptic ulcers,¹³ ulcerative colitis¹⁴ and dysplastic lesions in the colon¹⁵ with varying outcomes.

All endoscopists carrying out endoscopic retrograde cholangio-pancreaticography (ERCP) recognize the obvious differences in the appearance of the papilla of Vater. There are several descriptions of the papilla of Vater, anatomically,^{16–18} radiologically^{19,20} and histologically,²¹ but none of these contain any structured classification of the macroscopic appearance of the papilla of Vater as perceived during the endoscopic examination. Moreover, there is a widespread opinion that the macroscopic appearance might well co-vary with the cannulation complexity,^{22,23} complication rates²⁴ and also with other factors of relevance for the management of individual patients.²⁵ To comprehensively assess the association between the macroscopic appearance of the papilla and the cannulation complexity and other related clinical issues, a system is requested to accurately and reliably describe the papilla. The objectives of this study were therefore twofold; first to present an endoscopic classification of the papilla of Vater and secondly to determine inter- and intra-observer agreement among experts as well as non-expert endoscopists when assessing the different papillae types.

Material and methods

Still images from 140 different patients were captured during clinically indicated ERCP examinations, with standard duodenoscopes (TJF-Q180V, Olympus Medical Systems Co., Tokyo, Japan) connected to standard endoscopy processors (CV-180, Olympus).

No one of the patients had previously undergone treatment such as endoscopic sphincterotomy, papillary dilatation or biliary stenting. No gross pathology or anatomical variant, such as a juxtapapillary diverticula or tumour, was present. A set of at least four still images of the same papilla was captured before the clinical procedure started. First photographs were taken of the papilla and the surrounding duodenal wall at full inflation at two different angles, followed by one close-up view and finally one frontal image with a standard sphincterotome positioned on the side of the papilla, acting as a yardstick (Figure 1).

During the work on a previous study in the Scandinavian Association for Digestive Endoscopy (SADE) study group of ERCP concerning difficult cannulation,²⁶ the possible impact of the macroscopic appearances of the papilla were more closely discussed. With that discussion in mind, all image sets were presented to a group of expert endoscopists at our own institution, whereupon the papilla images were separated and collected into four different types based on their respective endoscopic appearance (Figure 1).

Type 1; Regular papilla, with no distinctive features i.e. 'classic appearance'.

Type 2; Small papilla, often flat, with a diameter not bigger than 3 mm (approximately 9 Fr).

Type 3; Protruding or pendulous papilla. A papilla is standing out, protruding or bulging into the duodenal lumen or sometimes hanging down, pendulous with the orifice oriented caudally.

Type 4; Creased or ridged papilla, where the ductal mucosa seems to extend distally, rather out of the papillary orifice, either on a ridge or in a crease.

To assess the accuracy by which different endoscopists could group the different images into the four types (interobserver agreement), a web-based survey was constructed containing 50 sets of still images of various papilla distributed between the four different types. Eighteen endoscopists were invited to participate in the survey, of whom nine were expert endoscopists from the Nordic countries and nine were non-experts recruited at random from participants in an introductory ERCP training course. The survey were constructed and distributed in a web-based interface (Analyzer Survey Solution, www.analyzer.se) accessible only to the invited endoscopists and investigators. No patient data were available, nor were patients' characteristics or symptoms presented. The survey started with an introduction of the predefined classification system followed by a few training exercises, with direct feedback to the respondents upon each decision. During the sharp survey the respondents were asked to select one out of the four standard types, which they

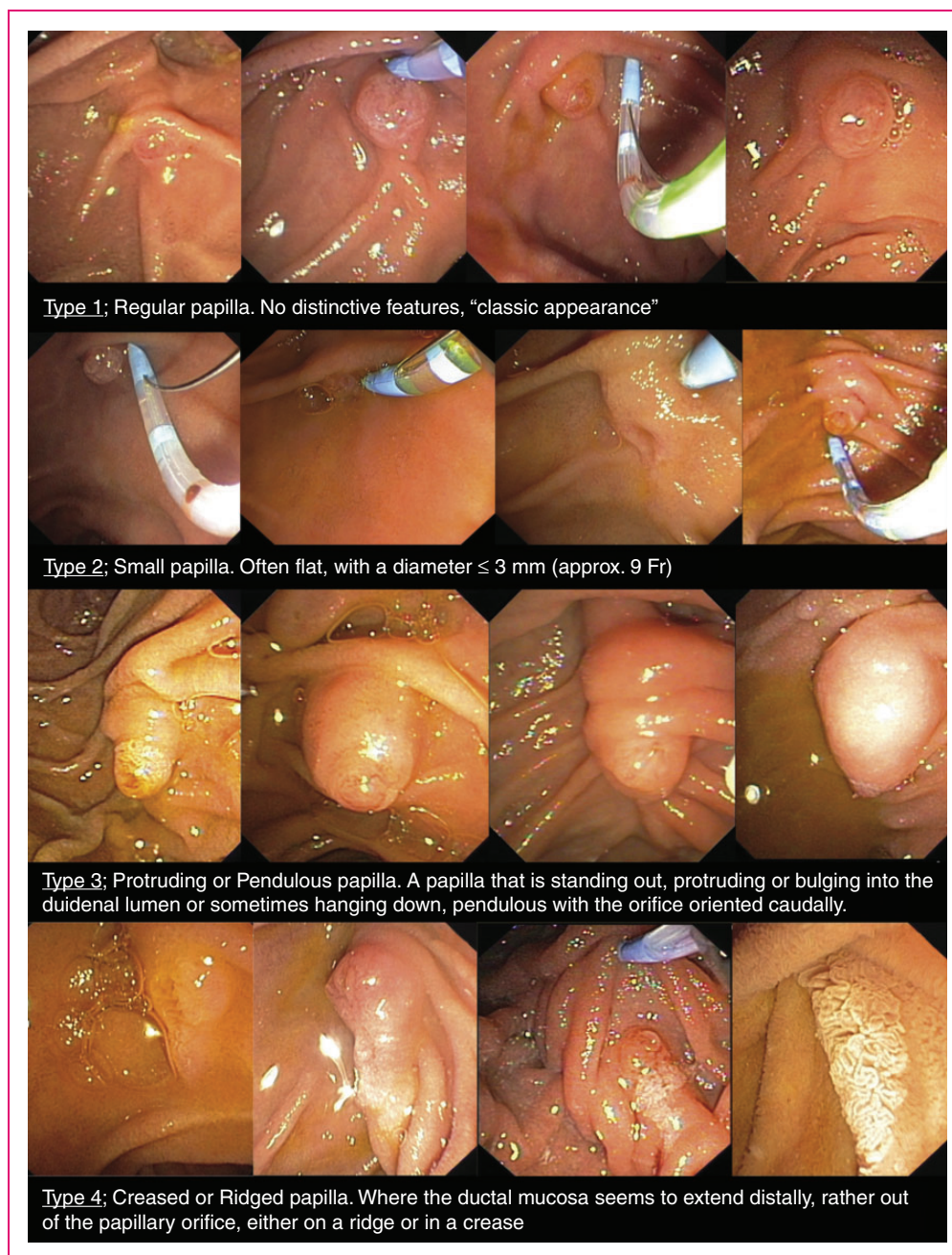


Figure 1. Endoscopic classification of the macroscopic appearance of the Papilla of Vater.

thought each individual set of images portrayed, by means of multiple choice assessment. The endoscopists had unlimited time to view each set of pictures, and were allowed to go back and change a previous decision. The survey presented in total 15 cases with a Type 1, 11 with Type 2, 13 with Type 3 and finally 11 cases with Type 4 papilla. We also collected information on the ERCP experience of the respective participant as well as the annual ERCP volume per institution. Data were anonymously collected and entered directly into a coded database.

Three months later the same endoscopists were asked to respond to a similar survey containing a stratified random sample of 20 sets (five each of the four different standard types) allowing for an assessment of intraobserver agreement.

Statistics and ethics

Interobserver agreement analysis was carried out for the entire group of respondents, after which a sub-analysis was done by dividing the respondents into

groups based on their respective level of ERCP experience. Likewise, intraobserver agreement was calculated by comparing the respondent's answers in both surveys.

The agreement between the respondents' decisions and the predefined classification was measured using kappa statistics. The agreement, as reflected by the kappa value, was interpreted as suggested by Landis and Koch²⁷ as either; slight ($\kappa \leq 0.2$), fair ($0.21 \leq \kappa \leq 0.40$), moderate ($0.41 \leq \kappa \leq 0.60$), substantial ($0.61 \leq \kappa \leq 0.80$) to almost perfect ($0.81 \leq \kappa \leq 1.00$).

The 95% confidence interval (CI) was calculated to measure the level of precision of the respective kappa values. The analysis was done using IBM SPSS Statistics, v. 23 (IBM, Chicago, Illinois, USA).

The regional ethics committee at Karolinska Institutet, Stockholm approved the study protocol (Dnr 2013/908-31/2).

Results

Nine expert endoscopists, and nine non-experts responded to the web-surveys (Table 1). Overall, the endoscopists were concordant with the predefined papilla type, in 72% (range 58–82%) of the cases. There were some variations between the different papilla types as to how often the endoscopists choice matched the predefined type. We noticed that in Type 3 and Type 4 there was a slightly higher level of agreement (Figure 2), while in Type 1 and Type 2 we observed a trend towards choosing to substitute one

type before the other. Overall we observed the interobserver agreement to be substantial for the entire group of endoscopists ($\kappa = 0.62$, 95% CI 0.59–0.66, Table 2) with similar results for both experts ($\kappa = 0.63$) and non-experts ($\kappa = 0.61$). The level of agreement between the individual endoscopists against the predefined classification varied from moderate ($\kappa = 0.44$) to substantial ($\kappa = 0.76$). Regarding intraobserver agreement, it also varied between individual endoscopists, from moderate ($\kappa = 0.50$) to almost perfect ($\kappa = 0.86$). The overall level of intraobserver agreement among the endoscopists showed substantial agreement ($\kappa = 0.66$, 95% CI 0.59–0.72) when the entire group was concerned. Again there was similar results among experts and non-experts (Table 2). To explore whether, the participants answered 'wrong' in both the first as well as in the second survey, thereby giving a false high intraobserver agreement, we compared the outcomes in the second survey with the predefined classification. This analysis again revealed a substantial agreement ($\kappa = 0.65$, 95% CI 0.59–0.71) with the predefined classification, harmonizing well with the outcome of the first survey. We were also, in these respects, unable to detect a difference between experts and non-experts.

Discussion

Many endoscopists recognize that there may be differences in the macroscopic visual appearance of the major duodenal papilla with potential to influence ERCP cannulation rates. In order to address this and related issues, a classification system has to be available which accurately describes the different types of papillae. Until now there has been no validated classification system to describe the endoscopic appearance of the papilla of Vater. Here we describe four different types of the virgin papilla of Vater and the validation process suggests that endoscopists can identify these four different gross appearances of the papilla with substantial interobserver as well as intraobserver agreement. Of importance also was our finding that we were unable to detect any differences between the expert and non-expert groups of endoscopists. These results illustrate the clinical usefulness of the classification system and that it has obvious potential to become an important research tool. Previous endoscopic classification systems for the papilla of Vater are scarce. Horiuchi et al.²⁸ described three different types of papillae, separated from each other depending on the degree of protrusion into the duodenal lumen. Their system was developed to act as a guide when choosing between different precut techniques in cases with so-called difficult cannulation. Based on Horiuchi et al.'s classification, Lee et al.²⁹ added a fourth 'distorted type', in their study regarding precut fistulotomy cannulation.

Table 1. Details the endoscopic retrograde cholangiopancreatography (ERCP) experience of the expert and the non-expert endoscopists.

	Experts (n = 9)	Non-experts (n = 9)
Years of ERCP experience		
Mean	20	2.5
Range	7–30	0–6
Annual number of ERCPs		
Mean	196	52
Range	100–400	25–75
Institutional ERCPs (cases/year)		
50–150		3
150–300		4
300–600	5	1
600 <	4	1
Lifetime ERCP experience (cases)		
<200		6
200–500		3
500–1500	1	
1500–	8	

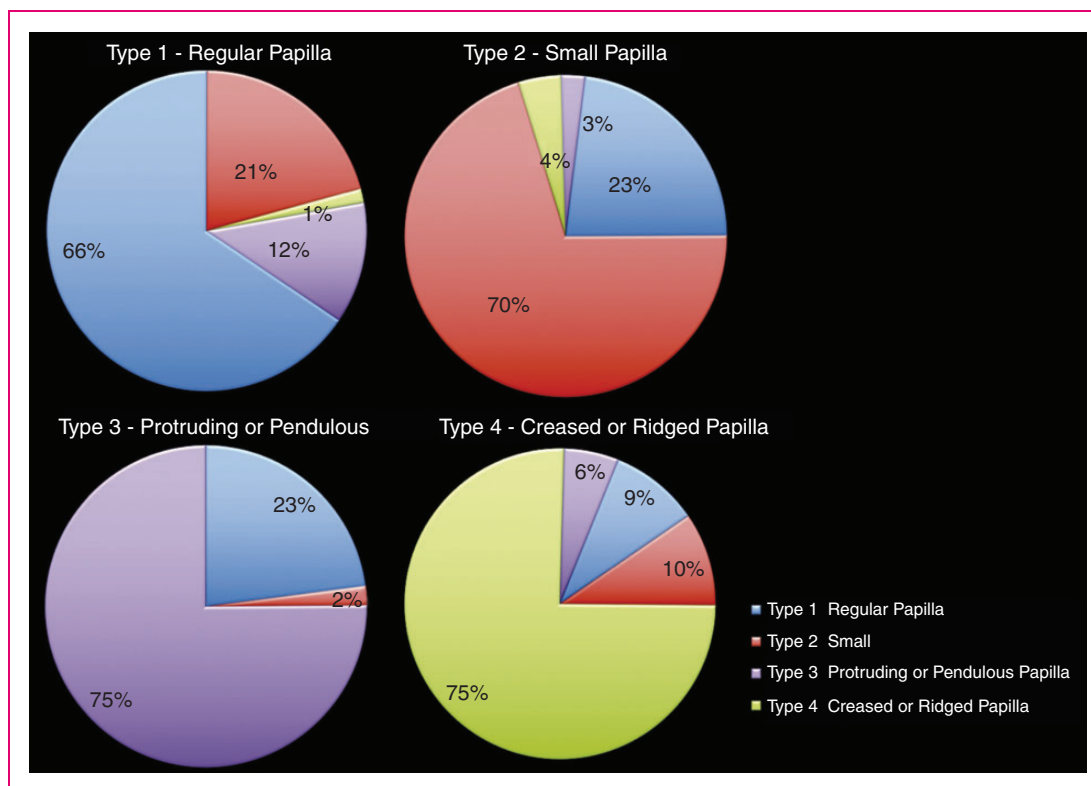


Figure 2. Observed agreement and distribution of alternative responses relative to the four predefined types of papillae.

Table 2. Interobserver agreement and intraobserver agreement among experts and non-experts.

	Observed agreement	Kappa (95% CI)	Agreement ²⁶
Interobserver agreement			
All respondents	0.72	0.62 (0.59–0.66)	Substantial
Experts	0.72	0.63 (0.57–0.69)	Substantial
Non-experts	0.72	0.61 (0.56–0.67)	Substantial
Intraobserver agreement			
All respondents	0.75	0.66 (0.59–0.72)	Substantial
Experts	0.77	0.68 (0.60–0.76)	Substantial
Non-experts	0.73	0.62 (0.53–0.72)	Substantial

CI: confidence interval.

However, it has to be remembered that none of these classification systems have been subject to a validation process, and neither of these studies include a description of a small or creased papilla. These studies do, however, emphasize the clinical relevance of an endoscopic classification of the papilla of Vater pertinent for the prediction of the complexity of the subsequent cannulation. Currently we have no data to offer on the cannulation success rates in the respective four types of papillae, but intuitively Types 2 and 3 are those in

which cannulation difficulties are foreseeable.^{24,25} Previous publications have determined technical aspects during cannulation to define ‘a papilla difficult to cannulate’³⁰ and subsequent studies have defined difficult cannulation,²⁶ criteria that now have been accepted to be used in future cannulation studies.³¹ These and related issues are currently specifically addressed in a forthcoming dedicated study dedicated study. Recent technical developments in capturing, storing and distributing digital images through digital media have paved the way for the use of high-quality image documentation, basically during all endoscopic procedures. These utilities will significantly improve both education and research, and allow endoscopic observations to be objectively verifiable and reproducible. Moreover, these digital and technological developments foster the processing of endoscopic classification systems for virtually the entire gastrointestinal tract. Unfortunately, studies as represented by the present investigation are often lagging behind, since the interpretation and reporting of endoscopic images may seriously suffer from precision and accuracy.³² Accordingly validation studies, such as inter- and intraobserver agreement evaluations, represent important tools in the attempt to standardize an otherwise ambiguous clinical and research situation.

The use of still images, instead of video sequences, can be criticized since the latter better mimics the clinical duodenoscopy situation. Previous studies in the field of endoscopic validation studies have, however, used still images of good quality and found these to be of sufficient value, not the least from a study logistic perspective.^{12,15} Video sequences are hampered by the fact that they are time-consuming to watch and evaluate and are often difficult to produce and reproduce with optimal image quality and proper freeze framing.

Furthermore, there exists some controversy regarding the use and relevance of kappa statistics. The calculated values in kappa statistics are influenced by the distribution of the traits in the data sets and are difficult to compare between different studies or study populations. Conclusions drawn from a single kappa value have to be made with caution, bearing in mind that that the value is just descriptive and needs to be interpreted. Accordingly, one single value is of limited or no value but the entire series of outcomes should be in focus for the reported outcomes. In the present study these outcomes were consistent throughout the various analytical exercises. Corresponding kappa analyses contain no component of hypothesis testing or comparisons between groups to distinguish if a value is 'true' or not. This makes power or sample size calculations obsolete but it has to be mentioned that the number of participants as currently practised are in accord with other studies in the field.³³

In conclusion, the proposed endoscopic classification of the papilla of Vater seems to be easy to use, irrespective of the level of experience of the endoscopist. It carries a substantial level of inter- and intraobserver agreement reflecting its potential to be clinically useful, and now the potential clinical relevance of the four different papilla types awaits to be determined.

Acknowledgements

The authors wish to express their thanks to Salmir Nasic for valuable statistical support. The SADE study group of ERCP members are: Lars Aabakken, Department of Cancer, Surgery and Transplant Surgery, Oslo University Hospital, Norway; Juha Grönroos, Division of Digestive Surgery and Urology, Turku University Hospital, Finland; Jorma Halttunen, Department of Gastrointestinal and General Surgery, Helsinki University Central Hospital, Finland; Truls Hauge, Department of Gastroenterology, Oslo University Hospital, Norway; Björn Lindkvist, Sahlgrenska Academy, University of Gothenburg, Sweden; Marja-Leena Kylänpää, Department of Gastrointestinal and General Surgery, Helsinki University Central Hospital, Finland; Pallo N Schmidt, Department of Gastroenterology and Gastrointestinal Surgery, Hvidovre Hospital, Denmark; Arto Saarela, Gastrointestinal Surgery Division, Oulu University Hospital, Finland; Ervin Toth, Department of

Clinical Sciences, Department of Gastroenterology and Nutrition, Skåne University Hospital, Lund University, Sweden.

Declaration of conflict of interests

The authors declare that there is no conflict of interest.

Funding

The study was supported by grants from the Research Fund at Skaraborg Hospital, Skövde (VGSKAS 308021) and from the Stockholm County Council, the Karolinska Institutet (SLL: ALF 20130512 to UA).

References

1. Aabakken L, Rembacken B, LeMoine O, et al. Minimal standard terminology for gastrointestinal endoscopy - MST 3.0. *Endoscopy* 2009; 41: 727–728.
2. Aabakken L, Barkun AN, Cotton PB, et al. Standardized endoscopic reporting. *J Gastroenterol Hepatol* 2014; 29: 234–240.
3. Löhr J-M, Lönnebro R, Stigliano S, et al. Outcome of probe-based confocal laser endomicroscopy (pCLE) during endoscopic retrograde cholangiopancreatography: A single-center prospective study in 45 patients. *United European Gastroenterol J* 2015; 3: 551–560.
4. Kamiński MF, Hassan C, Bisschops R, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2014; 46: 435–449.
5. Tringali A, Lemmers A, Meves V, et al. Intraductal biliopancreatic imaging: European Society of Gastrointestinal Endoscopy (ESGE) technology review. *Endoscopy* 2015; 47: 739–753.
6. Bendtsen F, Skovgaard LT, Sørensen TI, et al. Agreement among multiple observers on endoscopic diagnosis of esophageal varices before bleeding. *Hepatology* 1990; 11: 341–347.
7. Baldaque-Silva F, Marques M, Lunet N, et al. Endoscopic assessment and grading of Barrett's esophagus using magnification endoscopy and narrow band imaging: Impact of structured learning and experience on the accuracy of the Amsterdam classification system. *Scand J Gastroenterol* 2013; 48: 160–167.
8. Wallace MB, Sharma P, Lightdale C, et al. Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in Barrett's esophagus with probe-based confocal laser endomicroscopy. *Gastrointest Endosc* 2010; 72: 19–24.
9. Rath HC, Timmer A, Kunkel C, et al. Comparison of interobserver agreement for different scoring systems for reflux esophagitis: Impact of level of experience. *Gastrointest Endosc* 2004; 60: 44–49.
10. Lundell LR, Dent J, Bennett JR, et al. Endoscopic assessment of oesophagitis: Clinical and functional correlates and further validation of the Los Angeles classification. *Gut* 1999; 45: 172–180.

11. Hirano I, Moy N, Heckman MG, et al. Endoscopic assessment of the oesophageal features of eosinophilic oesophagitis: Validation of a novel classification and grading system. *Gut* 2013; 62: 489–495.
12. van Rhijn BD, Warners MJ, Curvers WL, et al. Evaluating the endoscopic reference score for eosinophilic esophagitis: Moderate to substantial intra- and interobserver reliability. *Endoscopy* 2014; 46: 1049–1055.
13. de Groot N, van Oijen M, Kessels K, et al. Reassessment of the predictive value of the Forrest classification for peptic ulcer rebleeding and mortality: Can classification be simplified? *Endoscopy* 2013; 46: 46–52.
14. de Lange T, Larsen S and Aabakken L. Inter-observer agreement in the assessment of endoscopic findings in ulcerative colitis. *BMC Gastroenterol* 2004; 4: 9.
15. Wanders LK, Mooiweer E, Wang J, et al. Low interobserver agreement among endoscopists in differentiating dysplastic from non-dysplastic lesions during inflammatory bowel disease colitis surveillance. *Scand J Gastroenterol* 2015; 50: 1011–1017.
16. Flati G, Flati D, Porowska B, et al. Surgical anatomy of the papilla of Vater and biliopancreatic ducts. *Am Surg* 1994; 60: 712–718.
17. Avisse C, Flament JB and Delattre JF. Ampulla of Vater. Anatomic, embryologic, and surgical aspects. *Surg Clin North Am* 2000; 80: 201–212.
18. Horiguchi S-I and Kamisawa T. Major duodenal papilla and its normal anatomy. *Dig Surg* 2010; 27: 90–93.
19. Guo Z-J, Chen Y-F, Zhang Y-H, et al. CT virtual endoscopy of the ampulla of Vater: Preliminary report. *Abdom Imaging* 2010; 36: 514–519.
20. Kim TU, Kim S, Lee JW, et al. Ampulla of Vater: Comprehensive anatomy, MR imaging of pathologic conditions, and correlation with endoscopy. *Eur J Radiol* 2008; 66: 48–64.
21. Paulsen FP, Bobka T, Tsokos M, et al. Functional anatomy of the papilla Vateri. *Surg Endosc* 2001; 16: 296–301.
22. Jamry A. Comparative analysis of endoscopic precut conventional and needle knife sphincterotomy. *World J Gastroenterol* 2013; 19: 2227–2233.
23. Bakman YG and Freeman ML. Difficult biliary access at ERCP. *Gastrointest Endosc Clin N Am* 2013; 23: 219–236.
24. Matsushita M, Uchida K, Nishio A, et al. Small papilla: Another risk factor for post-sphincterotomy perforation. *Endoscopy* 2008; 40: 875–876.
25. Williams EJ, Ogollah R, Thomas P, et al. What predicts failed cannulation and therapy at ERCP? Results of a large-scale multicenter analysis. *Endoscopy* 2012; 44: 674–683.
26. Halttunen J, Meisner S, Aabakken L, et al. Difficult cannulation as defined by a prospective study of the Scandinavian Association for Digestive Endoscopy (SADE) in 907 ERCPs. *Scand J Gastroenterol* 2014; 49: 752–758.
27. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
28. Horiuchi A, Nakayama Y, Kajiyama M, et al. Effect of precut sphincterotomy on biliary cannulation based on the characteristics of the major duodenal papilla. *Clin Gastroenterol Hepatol* 2007; 5: 1113–1118.
29. Lee TH, Bang BW, Park S-H, et al. Precut fistulotomy for difficult biliary cannulation: Is it a risky preference in relation to the experience of an endoscopist? *Dig Dis Sci* 2011; 56: 1896–1903.
30. Löhr JM, Aabakken L, Arnelo U, et al. How to cannulate? A survey of the Scandinavian Association for Digestive Endoscopy (SADE) in 141 endoscopists. *Scand J Gastroenterol* 2012; 47: 861–869.
31. Dumonceau J-M, Andriulli A, Elmunzer BJ, et al. Prophylaxis of post-ERCP pancreatitis: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - updated June 2014. *Endoscopy* 2014; 46: 799–815.
32. Mette Asfeldt A, Straume B and Paulssen EJ. Impact of observer variability on the usefulness of endoscopic images for the documentation of upper gastrointestinal endoscopy. *Scand J Gastroenterol* 2015; 42: 1106–1112.
33. Curvers W, Bohmer C, Mallant-Hent R, et al. Mucosal morphology in Barrett's esophagus: Interobserver agreement and role of narrow band imaging. *Endoscopy* 2008; 40: 799–805.